

Original citation:

Becker, Sascha O., Cinnirella, Francesco, Hornung, Erik and Woessmann, Ludger (2012) iPEHD – The Ifo Prussian Economic History Database. Working Paper. Coventry, UK: Department of Economics, University of Warwick. (CAGE Online Working Paper Series).

Permanent WRAP url:

<http://wrap.warwick.ac.uk/57890>

Copyright and reuse:

The Warwick Research Archive Portal (WRAP) makes this work of researchers of the University of Warwick available open access under the following conditions. Copyright © and all moral rights to the version of the paper presented here belong to the individual author(s) and/or other copyright owners. To the extent reasonable and practicable the material made available in WRAP has been checked for eligibility before being made available.

Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge. Provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

A note on versions:

The version presented here is a working paper or pre-print that may be later published elsewhere. If a published version is known of, the above WRAP url will contain details on finding it.

For more information, please contact the WRAP Team at: publicatons@warwick.ac.uk



<http://wrap.warwick.ac.uk/>

September 2012

No.96

**iPEHD –
The ifo Prussian Economic History Database**

Sascha O. Becker, Francesco Cinnirella, Erik Hornung and Ludger
Woessmann

WORKING PAPER SERIES

Centre for Competitive Advantage in the Global Economy

Department of Economics

iPEHD – The ifo Prussian Economic History Database*

Sascha O. Becker (U Warwick)[†]

Francesco Cinnirella (Ifo)

Erik Hornung (Ifo)

Ludger Woessmann (U Munich and Ifo)

This paper provides a documentation of the ifo Prussian Economic History Database (iPEHD), a county-level database covering a rich collection of variables for 19th-century Prussia. The Royal Prussian Statistical Office collected these data in several censuses over the years 1816-1901, with much county-level information surviving in archives. These data provide a unique source for micro-regional empirical research in economic history, enabling analyses of the importance of such factors as education, religion, fertility, and many others for Prussian economic development in the 19th century. The service of iPEHD is to provide the data in a digitized and structured way.

Keywords: economic history, Prussia, 19th century, database, county

JEL classification: N13, N33

June 21, 2012

* Over the years, a large number of research assistants have contributed to the digitization work underlying iPEHD. We are grateful for their contributions, especially to Christian Steibl, as well as to Rajesh Bhateja. Financial support by the Pact for Research and Innovation of the Leibniz Association is gratefully acknowledged.

[†] Becker: Department of Economics and CAGE, University of Warwick, Coventry, CV4 7AL, United Kingdom; also Ifo, CEPR, CESifo, and IZA; s.o.becker@warwick.ac.uk. Cinnirella: Ifo Institute, Poschingerstr. 5, 81679 Munich, Germany; also CESifo and CEPR; cinnirella@ifo.de. Hornung: Ifo Institute, Poschingerstr. 5, 81679 Munich, Germany; hornung@ifo.de. Woessmann: University of Munich and Ifo Institute, Poschingerstr. 5, 81679 Munich, Germany; also CESifo, IZA, and CAGE; woessmann@ifo.de.

1. Introduction

Prussian economic history during the 19th century proves a fascinating setting to study many of the most fundamental questions in economic history. A country of such high diversity, but with a rather uniform institutional setting, allows answering many important research questions by analyzing the micro-regional variation existing within one country. For example, the Prussian setting allows analyzing the importance of such factors as education, religion, fertility, and many others for industrialization and historical economic development.

What is more, starting with the first full-scale population census in 1816, the Royal Prussian Statistical Office collected a huge amount of data in a number of censuses over the 19th century. Many interesting county-level data have survived in archives. Thanks to the Prussian proverbial orderliness and thoroughness, we have high-quality data for the Prussian counties (*Kreise*) covering nearly the whole range of the 19th century. These data provide a unique source for empirical research in economic history, with the particular potential to study historical micro-regional data with modern microeconomic methods.

The service of the ifo Prussian Economic History Database (iPEHD) is to provide many of these data in a digitized and structured way. Thus, iPEHD is a county-level database covering a rich collection of variables for all counties of Prussia over the period 1816-1901. iPEHD is freely available to any interested researcher at www.cesifo-group.de/ipehd. The iPEHD website does not only provide the raw data, but also background information, definitions, and sources of variables. It also makes suggestions on how to merge data from different census waves with varying administrative boundaries into panel datasets. Finally, it provides a collection of thematic maps visualizing the data, ready-made datasets and codes to replicate tables from published research, and additional material.

Throughout, iPEHD covers all Prussian counties, whose number increased over the 19th century from 308 in 1816 to 574 in 1901. Drawing from a total of 15 original sources – many of which contain several volumes – iPEHD comprises more than 1,500 variables. The available data cover a wide range of topics including a host of indicators of economic development, education, demographics, and more. iPEHD organizes these data into eight content areas: education, occupation, wages and income tax, industry, agriculture, population, religion, and miscellaneous. In total, iPEHD contains more than half a million data points at the county level. While nowhere near being a complete collection of all available data, we think that iPEHD provides a comprehensive micro-regional database on 19th-century economic history in Prussia.

This paper documents iPEHD and provides guidance on how to use the data contained in it. The next section starts with some brief background on how iPEHD emerged. Section 3 provides an overview of the data contained in iPEHD. Section 4 describes the data structure and suggests a procedure to combine data from different census years. Section 5 lists the original sources, published by the Royal Prussian Statistical Bureau or its employees, from which the iPEHD data stem. Section 6 gives a brief overview of research that has been conducted using iPEHD data so far. Section 7 presents a few additional features of the iPEHD website, and Section 8 concludes.

2. A Brief History of the Birth of iPEHD

In 2006, when looking for data to analyze the relationship of religion and literacy with economic outcomes in German history, we stumbled upon the rich county-level data available from the Prussian census of 1871. After thorough studies of the data, we were fascinated by the depth and breadth of the historical information that the Royal Prussian Statistical Office had collected and documented. Prussian thoroughness had produced high-quality data at the county level in the 19th century documenting everything from education over religion and demographics to economic development (see Figure 1 for an example).

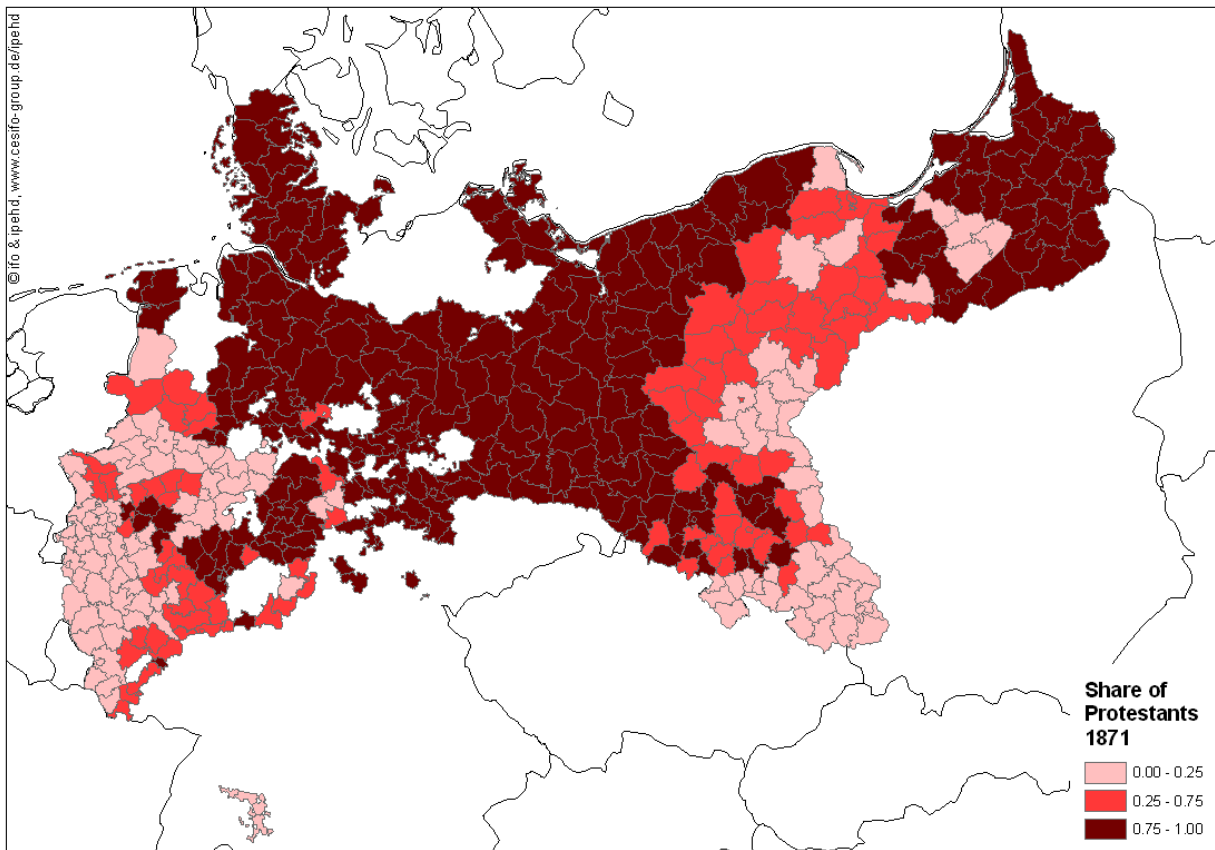
Soon, we recognized the sheer amount of data that were just sitting around in the statistical annals at German state libraries. The quality of this impressive collection of information, remarkable for the 19th century, has generally been regarded as excellent by historians and demographers (cf., e.g., Galloway, Hammel, and Lee (1994)).¹ And compared to the selective samples which a lot of historical research is restricted to, the full censuses covering the whole population provide a much more reliable picture of the historical setting.

After the original “Was Weber Wrong?” paper (eventually published as Becker and Woessmann (2009)) which relied mainly on the 1871 census and subsequent data, we explored annals covering rather unknown census data from 1816 to 1821.² Although lots of effort had to be undertaken to make these data ready for research and to ensure their comparability, we soon found it to be very promising and equally reliable. A third large data digitization project involved the census of 1849. The sheer amount of information provided in the sources was overwhelming.

¹ After we had digitized the data used in Becker and Woessmann (2009), data from that project became available online at www.patrickgalloway.com.

² We are grateful to Davide Cantoni for pointing us to these data sources at the time.

Figure 1: Protestantism in 19th-Century Prussia



Note: County-level depiction based on the 1871 Prussian Population Census. For details, see Becker and Woessmann (2009).

The censuses of 1816, 1849, and 1871 became the foundation of iPEHD. But, as time went by, we also digitized data from different other censuses to fill in the gaps. Although far from complete, we find the data to provide a rather comprehensive overview of 19th-century economic history in Prussia. Thus, we are happy to be able to make the digitized data available to the scientific community and the interested public. iPEHD went online in the summer of 2012 to be freely used by anyone interested at www.cesifo-group.de/ipehd.

The collection of these data and their provision to the scientific community is part of the project “Establishment of a leading international center for empirical research on the importance of education for long-term economic development,” generously funded by the Leibniz Association under the Pact for Research and Innovation. The project was carried out at the Human Capital and Innovation department at the Ifo Institute – Leibniz Institute for Economic Research at the University of Munich.

3. An Overview of the Data contained in iPEHD

This section provides an overview of iPEHD, discussing its scope, the structure of its data files, the areas of content covered, and the information contained in the codebooks.

3.1 *The Scope and Data Files of iPEHD*

iPEHD starts with the population census in 1816, the first full-scale census released by the Royal Prussian Statistical Office, which had been founded in 1805. The 1816 census covers the 308 Prussian counties at the time. Further extensive census data are available in 1849, 1864, 1871, and 1882, but – as indicated in Table 1 – many more detailed data were collected in additional years. As the number of counties grew over time, by 1901 the data cover 574 Prussian counties. In total, iPEHD contains more than 1,500 variables and more than half a million data points, all at the county level.

Table 1: The Scope of iPEHD

Year	No. of variables	No. of county observations	No. of data points
1816	58	308	17,864
1819	5	344	1,720
1821	22	344	7,568
1816-1821	24	456	10,944
1829	6	59	354
1849	712	335	238,520
1858	6	342	2,052
1862	4	346	1,384
1864	53	347	18,391
1866a	1	342	342
1866b	11	334	3,674
1871a	25	453	11,325
1871b	14	458	6,412
1878	5	426	2,130
1882a	269	464	124,816
1882b	14	465	6,510
1886a	156	544	84,864
1886b	97	518	50,246
1892	8	550	4,400
1896	15	552	8,280
1901	8	574	4,592
Sum	1,513		606,388

Note: Some of the data points may contain missing information.

iPEHD consists of county-level information gathered from these different censuses. The data are currently presented in 76 separate data files, organized by content area, specific topic, and census year. Each data file in iPEHD contains a unique county (*Kreis*) identifier (discussed in Section 4.2 below), the county name, the abbreviated district (*Regierungsbezirk*) name (rb), and a set of variables of census data. iPEHD stores its data in comma-separated values (csv) format, which is easily accessible from any statistical software. For example, to open the csv data files in Stata, one just has to type:

```
insheet using "xxxxxx.csv"
```

To give an example of a data file, Table 2 shows a brief extract of a few variables for the first few counties (by alphabet) from the data file “ipehd_1819_indu_fac.csv”, which contains data on the number of factories in a county in 1819. E.g., the variable “fac1819_brick” documents the total number of brick manufactories in a county in 1819, and the variable “mill1819_water” the total number of water mills.

Table 2: Extract from an Example Data File

kreiskey1800	County	rb	fac1819_brick	fac1819_lime	fac1819_glass	mill1819_water
277	Achen	AAC	5	10	2	26
33	Adelnau	POS	11	6	0	26
254	Adenau	KOB	0	1	0	71
196	Ahaus	MUN	11	15	0	20
255	Ahrweiler	KOB	0	0	0	51
2	Allenstein	KON	5	0	1	31
219	Altena	ARN	3	13	0	41
257	Altenkirchen	KOB	1	0	0	41
10	Angerburg	GUM	4	26	0	5
53	Angermünde	POT	13	2	0	28
32	Anklam	STE	3	0	0	2
209	Arnsberg	ARN	12	4	0	26
67	Arnswalde	FRA	7	3	3	29
160	Aschersleben	MAG	8	5	0	57
55	(Nieder-)Barnim	POT	8	0	1	30
54	(Ober-)Barnim	POT	18	0	0	36
190	Beckum	MUN	8	3	0	22

Note: Extract from iPEHD data file “ipehd_1819_indu_fac.csv”.

3.2 Areas of Content covered by iPEHD

The iPEHD data are categorized into the following eight content areas:

Education: This area contains, among others, such data as the number of students, teachers, and schools by school type, literacy, and school finance.

Occupation: This area contains, among others, data on the labor force in agriculture, in factories, in manufacturing, in crafts, and in services.

Wages and Income Tax: This area contains data on daily wages of day laborers, on teacher income, and on income taxes.

Industry: This area contains data on a huge number of different factories, technologies, and transportation.

Agriculture: This area contains, among others, such data as livestock, crop yields, soil composition, and the distribution of land.

Population: This area contains data on population by age, by gender, and by marital status, on births and deaths, and on population with disabilities.

Religion: This area contains denomination-specific data on population, literacy, education, occupation, and number of churches.

Miscellaneous: This area contains such data as surface area, buildings, municipalities, and residential areas for each county.

Apart from the data gathered in these eight content areas, the **merger file** provides information on merger variables necessary to combine data from different census years (see Section 4.3 below).

3.3 Codebooks

A large number of codebooks provide additional information for each variable contained in iPEHD. There is one codebook for each year, so that explanations for each variable can be found in the codebook for the corresponding year. A summary codebook that combines all years is also provided; this summary codebook allows a content search of the whole iPEHD.

The codebooks list the variable name (“variable name”), the name of the data file where it can be found (“ipehd datasets”), an English label (“label”), and the original label in German language (“original label”). The German language label is similar to the table headings found in the original sources. The English label leads with the year and is a shortened (direct) translation

of the German label; in cases where a translation is not feasible, the original German term was adopted. In addition, the codebooks indicate the source of each set of variables (“source”).

4. Merging Data from Different Censuses

One of the biggest challenges when analyzing historical data is to ensure comparability over time. A key service of iPEHD is to facilitate the analysis of data from different census years at the county level, holding the administrative boundaries fixed. This section presents the structure in which the data are presented in iPEHD and the suggested procedure of combining different census years.

4.1 County-level Structure of the Data

Starting after the Congress of Vienna in 1815, Prussia reformed its administrative structure and introduced the county level. At the time, the dimension of a county was meant to follow borders of previously existing administrative units. The maximum distance to the administrative center was meant to be two to three Prussian Miles (roughly 15 to 23 km or 9 to 14 miles), such that every inhabitant could travel forth and back within a day. The population size was meant to range between 20,000 in sparsely populated areas and 36,000 in densely populated areas.

Throughout the 19th century, various administrative reforms reshaped the county structure of Prussia. As the population grew over time, it became necessary to divide existing administrative units in order to reduce administrative efforts. Most of these changes were partitions of one county into two or more counties.

Thus, it is usually possible to reconstruct earlier administrative units by aggregating data from later years to the former structure. A drawback of this procedure is that the researcher loses part of the variation provided by having more observations. Still, the procedure appears necessary in order to have intertemporal comparability of the units of observation. The alternative would be to assign the same early data to two or more subsequently parted units, introducing measurement error if observed data were not uniformly distributed in the original area.

A peculiarity of the Prussian county system is the city county. Starting with the introduction of the county level in 1815, the so-called *Immediatstädte* (immediate towns) became a county themselves. As urbanization advanced, an increasing number of cities were detached from their

original county and became a county of their own. Thus, the database often contains a *Landkreis* (rural county) and a *Stadtkreis* (city county) with similar names. For example, there are six pairs of *Landkreis/Stadtkreis* information among the 335 county observations in the 1849 classification and 20 pairs among the 458 county observations in the 1874 classification.

4.2 County Identifiers

All data in iPEHD reflect the administrative conditions in place at the date of publication of the census. Since censuses often ordered the counties in different ways, identifiers were assigned reflecting the order of each census. Thus, each county in each census has been assigned a continuous number which is *unique within a census but not across censuses*. The identifiers are named kreiskeyYYYY, where YYYY represents the four-digit year (see Section 4.5 below for additional peculiarities of the 1816-21 data).

The year in the identifier denotes the administrative structure of Prussia, which is not necessarily the same as the census year. In some cases, different identifiers (e.g., kreiskey1871 and kreiskey1874) even had to be assigned to data from the same census year (1871) because the Royal Prussian Statistical Office used different aggregations in different publications of data from the same census.

4.3 Intertemporal Comparisons

Researchers may be interested in intertemporal comparisons and in the construction of panel datasets using iPEHD. The iPEHD county identifiers, together with the merge-county file also provided on the iPEHD website, provide a service that facilitates such linkage of comparable units of observation over time. Our suggestion is that, *in order to obtain a comparable set of observations, researchers should collapse the data to the earliest set of counties in the data*. However, it is important to point out that at the end of the day, the best way to structure and use the data will be specific to every single research project.

To conduct intertemporal comparisons, our suggestion is to take the following nine-step procedure. To construct cross-sections, the procedure should be followed only until step 3.

1. Choose datasets from the same census year.
2. Merge all datasets from the same census year using the identifier (e.g., kreiskey1882).
3. Save the cross-section.

4. Use the merge-county file provided on the iPEHD website.
5. Drop all duplicate and missing observations from the merge-county file according to the identifier in the cross-section (e.g., kreiskey1882): see section 4.4 for an example.
6. Merge the merge-county file with the cross-section using the identifier (e.g., kreiskey1882).
7. Aggregate (sum/mean) all variables in the cross-section to the aggregation level of the earliest census in the analysis using the identifier of the earliest census in the analysis (crucial step!).
8. Repeat steps 1 to 7 for datasets from other census years.
9. Merge the resulting cross-sections using the identifier of the earliest census in the analysis.

4.4 Example from the Merger File

In the example shown in Table 3, the eight illustrative counties observed in 1901 were established from six counties in 1874 and five counties in 1849. Between 1849 and 1874, the ‘Elbing Landkreis’ had been divided into ‘Elbing Stadtkreis’ and ‘Elbing Landkreis’. Between 1874 and 1901, the ‘Danzig Landkreis’ had been divided into ‘Danzig Niederung’, ‘Danzig Höhe’, and ‘Dirschau’.

Table 3: Example from the Merge File

Kreiskey 1901	County1901	Kreiskey 1874	County1874	Kreiskey 1849	County1849
38	ELBING STADTKREIS	38	ELBING STADTKREIS	37	ELBING LANDKREIS
39	ELBING LANDKREIS	39	ELBING LANDKREIS	37	ELBING LANDKREIS
40	MARIENBURG IN PREUSSEN	40	MARIENBURG IN PREUSSEN	38	MARIENBURG IN PREUSSEN
41	DANZIG STADTKREIS	41	DANZIG STADTKREIS	39	DANZIG STADTKREIS
42	DANZIG NIEDERUNG	42	DANZIG LANDKREIS	40	DANZIG LANDKREIS
43	DANZIG HOHE	42	DANZIG LANDKREIS	40	DANZIG LANDKREIS
44	DIRSCHAU	42	DANZIG LANDKREIS	40	DANZIG LANDKREIS
45	PREUSSISCH STARGARD	43	PREUSSISCH STARGARD	41	PREUSSISCH STARGARD

Note: Extract from the iPEHD merge file “ipehd_merge_county.csv”.

In order to have a comparable set of observations when performing intertemporal comparisons between 1901 and 1849, one has to aggregate the observations of ‘Danzig Niederung’, ‘Danzig Höhe’, and ‘Dirschau’ to match ‘Danzig Landkreis’. Thus, one should always aggregate the data to the aggregation level of the earliest census year in the specific analysis (step 7).

However, to perform intertemporal comparisons between, e.g., 1874 and 1849, one needs to drop the duplicate entries of ‘Danzig Landkreis’ from the merger file first (step 5). In addition, one needs to drop entries from the merger file that have missing observations on the county identifier in the respective year. Such missing observations exist because some territories were annexed by Prussia only after the respective census year.

As one example of how to merge datasets from 1874 and 1849, the following Stata code exemplifies the nine steps of the suggested procedure:

```
insheet using "ipehd_1871_edu_literacy_part2.csv", clear    /* Step 1 */
save "ipehd_1871_edu_literacy_part2.dta"

insheet using "ipehd_1871_pop_demo_part2.csv", clear
save "ipehd_1871_pop_demo_part2.dta"

merge 1:1 kreiskey1874 using "ipehd_1871_edu_literacy_part2.dta"  /* Step 2 */
drop _merge

save "ipehd_1871_part2.dta"    /* Step 3 */

insheet using "ipehd_merge_county.csv", clear    /* Step 4 */

duplicates drop kreiskey1874, force    /* Step 5 */
drop if kreiskey1874==.

merge 1:1 kreiskey1874 using "ipehd_1871_part2.dta"    /* Step 6 */

collapse (sum) pop* lit* edu*, by (kreiskey1849)    /* Step 7 */
drop if kreiskey1849==.
save "ipehd_1871_part2_2.dta"

insheet using "ipehd_1849_rel_deno.csv", clear    /* Step 8 */
save "ipehd_1849_rel_deno.dta"

insheet using "ipehd_merge_county.csv", clear
duplicates drop kreiskey1849, force
drop if kreiskey1849==.
merge 1:1 kreiskey1849 using "ipehd_1849_rel_deno.dta"
collapse (sum) rel*, by (kreiskey1849)
save "ipehd_1849.dta"

merge 1:1 kreiskey1849 using "ipehd_1871_part2_2.dta"    /* Step 9 */
drop _merge
save "ipehd_1849_1871.dta"
```

4.5 Peculiarity of the Data from 1816 to 1821

By 1816, Prussia had just started her administrative reform that established the county level. In some parts of the country, the reforms had not been finalized even in 1821. Thus, the data from the censuses in 1816 until after 1821 sometimes reflect old administrative units.

Unfortunately, due to the reform, these old units were subsequently aggregated and then newly divided in order to establish the new counties. This makes it impossible to accurately match the data of (some of) the administrative units from the early censuses to (some) counties in subsequent censuses. The *kreiskey1800* is thus coded so as to aggregate the data to a higher level. The *kreiskey1800* can be used to link the 1816-1821 data to later periods.

However, iPEHD also provides a unique identifier that allows merging data from the same census for these cross-sections. These identifiers are named 'id1816' and 'id1819'. In order to merge data from 1816 to other data from 1816, one should use id1816. In order to merge data from 1819 or 1821 to other data from 1819 or 1821, one should use id1819.

In order to merge data from 1816, 1819, or 1821 to data from subsequent censuses, one should take the following steps:

1. Choose datasets from 1816, 1819, or 1821.
2. Merge all datasets from the same census using the identifier (idYYYY).
3. Aggregate (sum/mean) all cross-sections using the identifier 'kreiskey1800'.
4. Merge the cross-section with aggregated data from subsequent censuses using the identifier 'kreiskey1800'.

5. Original Sources underlying the iPEHD Data

The iPEHD data have been digitized from different sources originally published by the Royal Prussian Statistical Bureau or its employees. These original historical volumes should be consulted for detailed information regarding the exact attributes of the data. Figure 2 shows two example pages from such source volumes.

The following list documents all the volumes used as sources for iPEHD. There are a total of 15 original sources, many of which consist of several volumes.

1816-21: Mützell, Alexander A. (1821-25). *Neues Topographisch-statistisch-geographisches Wörterbuch des Preussischen Staats*, Vol. 1-6. Halle: Karl August Kümmel.

Figure 2: Example Pages from Source Volumes

234

Zusammenstellung der Kreise

Gemeinschaften, Landkreise und Kreise

235

und ihrer Bevölkerung

Ortsanwaltschaften am 1. Dezember 1871

Religion

Alter u. Schulbildung

Frei Geborene

Einwohner

1871

1870

1869

1868

1867

1866

1865

1864

1863

1862

1861

1860

1859

1858

1857

1856

1855

1854

1853

1852

1851

1850

1849

1848

1847

1846

1845

1844

1843

1842

1841

1840

1839

1838

1837

1836

1835

1834

1833

1832

1831

1830

1829

1828

1827

1826

1825

1824

1823

1822

1821

1820

1819

1818

1817

1816

1815

1814

1813

1812

1811

1810

1809

1808

1807

1806

1805

1804

1803

1802

1801

1800

1799

1798

1797

1796

1795

1794

1793

1792

1791

1790

1789

1788

1787

1786

1785

1784

1783

1782

1781

1780

1779

1778

1777

1776

1775

1774

1773

1772

1771

1770

1769

1768

1767

1766

1765

1764

1763

1762

1761

1760

1759

1758

1757

1756

1755

1754

1753

1752

1751

1750

1749

1748

1747

1746

1745

1744

1743

1742

1741

1740

1739

1738

1737

1736

1735

1734

1733

1732

1731

1730

1729

1728

1727

1726

1725

1724

1723

1722

1721

1720

1719

1718

1717

1716

1715

1714

1713

1712

1711

1710

1709

1708

1707

1706

1705

1704

1703

1702

1701

1700

1699

1698

1697

1696

1695

1694

1693

1692

1691

1690

1689

1688

1687

1686

1685

1684

1683

1682

1681

1680

1679

1678

1677

1676

1675

1674

1673

1672

1671

1670

1669

1668

1667

1666

1665

1664

1663

1662

1661

1660

1659

1658

1657

1656

1655

1654

1653

1652

1651

1650

1649

1648

1647

1646

1645

1644

1643

1642

1641

1640

1639

1638

1637

1636

1635

1634

1633

1632

1631

1630

1629

1628

1627

1626

1625

1624

1623

1622

1621

1620

1619

1618

1617

1616

1615

1614

1613

1612

1611

1610

1609

1608

1607

1606

1605

1604

1603

1602

1601

1600

1599

1598

1597

1596

1595

1594

1593

1592

1591

1590

1589

1588

1587

1586

1585

1584

1583

1582

1581

1580

1579

1578

1577

1576

1575

1574

1573

1572

1571

1570

1569

1568

1567

1566

1565

1564

1563

1562

1561

1560

1559

1558

1557

1556

1555

1554

1553

1552

1551

1550

1549

1548

1547

1546

1545

1544

1543

1542

1541

1540

1539

1538

1537

1536

1535

1534

1533

1532

1531

1530

1529

1528

1527

1526

1525

1524

1523

1522

1521

1520

1519

1518

1517

1516

1515

1514

1513

1512

1511

1510

1509

1508

1507

1506

1505

1504

1503

1502

1501

1500

1499

1498

1497

1496

1495

1494

1493

1492

1491

1490

1489

1488

1487

1486

1485

1484

1483

1482

1481

1480

1479

1478

1477

1476

1475

1474

1473

1472

1471

1470

1469

1468

1467

1466

1465

1464

1463

1462

1461

1460

1459

1458

1457

1456

1455

1454

1453

1452

1451

1450

1449

1448

1447

1446

1445

1444

1443

1442

1441

1440

1439

1438

1437

1436

1435

1434

1433

1432

1431

1430

1429

1428

1427

1426

1425

1424

1423

1422

1421

1420

1419

1418

1417

1416

1415

1414

1413

1412

1411

1410

1409

1408

1407

1406

1405

1404

1403

1402

1401

1400

1399

1398

1397

1396

1395

1394

1393

1392

1391

1390

1389

1388

1387

1386

1385

1384

1383

1382

1381

1380

1379

1378

1377

1376

1375

1374

1373

1372

1371

1370

1369

1368

1367

1366

1365

1364

1363

1362

1361

1360

1359

1358

1357

1356

1355

1354

1353

1352

1351

1350

1349

1348

1347

1346

1345

1344

1343

1342

1341

1340

1339

1338

1337

1336

1335

1334

1333

1332

1331

1330

1329

1328

1327

1326

1325

1324

1323

1322

1321

1320

1319

1318

1317

1316

1315

1314

1313

1312

1311

1310

1309

1308

1307

1306

1305

1304

1303

1302

1301

1300

1299

1298

1297

1296

1295

1294

1293

1292

1291

1290

1289

1288

1287

1286

1285

1284

1283

1282

1281

1280

1279

1278

1277

1276

1275

1274

1273

1272

1271

1270

1269

1268

1267

1266

1265

1264

1263

1262

1261

1260

1259

1258

1257

1256

1255

1254

1253

1252

1251

1250

1249

1248

1247

1246

1245

1244

1243

1242

1241

1240

1239

1238

1237

1236

1235

1234

1233

1232

1231

1230

1229

1228

1227

1226

1225

1224

1223

1222

1221

1220

1219

1218

1217

1216

1215

1214

1213

1212

1211

1210

1209

1208

1207

1206

1205

1204

1203

1202

1201

1200

1199

1198

1197

1196

1195

1194

1193

1192

1191

1190

1189

1188

1187

1186

1185

1184

1183

1182

1181

1180

1179

1178

1177

1176

1175

1174

1173

1172

1171

1170

1169

1168

1167

1166

1165

1164

1163

1162

1161

1160

1159

1158

1157

1156

1155

1154

1153

1152

1151

1150

1149

1148

1147

1146

1145

1144

1143

1142

1141

1140

1139

1138

1137

1136

1135

1134

1133

1132

1131

1130

1129

1128

1127

1126

1125

1124

1123

1122

1121

1120

1119

1118

1117

1116

1115

1114

1113

1112

1111

1110

1109

1108

1107

1106

1105

1104

1103

1102

1101

1100

1099

1098

1097

1096

1095

1094

1093

1092

1091

1090

1089

1088

1087

1086

1085

1084

1083

1082

1081

1080

1079

1078

1077

1076

1075

1074

1073

1072

1071

1070

1069

1068

1067

1066

1065

1064

1063

1062

1061

1060

1059

1058

1057

1056

1055

1054

1053

1052

1051

1050

1049

1048

1047

1046

1045

1044

1043

1042

1041

1040

1039

1038

1037

1036

1035

1034

1033

1032

1031

1030

1029

1028

1027

1026

1025

1024

1023

1022

1021

1020

1019

1018

1017

1016

1015

1014

1013

1012

1011

1010

1009

1008

1007

1006

1005

1004

1003

1002

1001

1000

999

998

997

996

995

994

993

992

991

990

989

988

987

986

985

984

983

982

981

980

979

978

977

976

975

974

973

972

971

970

969

968

967

966

965

964

963

962

961

960

959

958

957

956

955

954

953

952

951

950

949

948

947

946

945

944

943

942

941

940

939

938

937

936

935

934

933

932

931

930

929

928

927

926

925

924

923

922

921

920

919

918

917

916

915

914

913

912

911

910

909

908

907

906

905

904

903

902

901

900

899

898

897

896

895

894

893

892

891

890

889

888

887

886

885

884

883

882

881

880

879

878

877

876

875

874

873

872

871

870

869

868

867

866

865

864

863

862

861

860

859

858

857

856

855

854

853

852

851

850

849

848

847

846

845

844

843

842

841

840

839

838

837

836

835

834

833

832

831

830

829

828

827

826

825

824

823

822

821

820

819

818

817

816

815

814

813

812

811

810

809

808

807

806

805

804

803

802

801

800

799

798

797

796

795

794

793

792

791

790

789

788

787

786

785

784

783

782

781

780

779

778

777

776

775

774

773

772

771

770

769

768

767

766

765

764

763

762

761

760

759

758

757

756

755

754

753

752

751

750

749

748

747

746

745

744

743

742

741

740

739

738

737

736

735

734

733

732

731

730

729

728

727

726

725

724

723

722

721

720

719

718

717

716

715

714

713

712

711

710

709

708

707

706

705

704

703

702

701

700

699

698

697

696

695

694

693

692

691

690

689

688

687

686

685

684

683

682

681

680

679

678

677

676

675

674

673

672

671

670

669

668

667

666

665

664

663

662

661

660

659

658

657

656

655

654

653

652

651

650

649

648

647

646

645

644

643

642

641

640

639

638

637

636

635

634

633

632

631

630

629

628

627

626

625

624

623

622

621

620

619

618

617

616

615

614

1829: Preussisches Statistisches Landesamt (1829). *Beiträge zur Statistik der Königlichen Preussischen Rheinlande, aus amtlichen Nachrichten zusammengestellt*. Aachen: J.A. Mayer.

1849: Statistisches Bureau zu Berlin (1851-55). *Tabellen und amtliche Nachrichten über den Preussischen Staat für das Jahr 1849*, Vol. 1-6b. Berlin: Statistisches Bureau zu Berlin.

1858: Meitzen, August (1868). *Der Boden und die landwirthschaftlichen Verhältnisse des Preussischen Staates*, Vol. 1-4. Berlin: Verlag von Paul Parey.

1862: Königlich Preussisches Statistisches Bureau (1863). Die Eisen-, Stein- und Wasserstrassen des preussischen Staates im Jahre 1862, in *Zeitschrift des Königlich Preussischen Statistischen Bureaus*, Vol. 3, 206-214. Berlin: Verlag der Königlichen Geheimen Ober-Hofbuchdruckerei.

1864: Königliches Statistisches Bureau in Berlin (1867). *Die Ergebnisse der Volkszählung und Volksbeschreibung, der Gebäude und Viehzählung, nach den Aufnahmen vom 3. December 1864, resp. Anfang 1865 und die Statistik der Bewegung der Bevölkerung in den Jahren 1862, 1863 und 1864*. Preussische Statistik Vol. 10. Berlin: Verlag von Ernst Kuehn.

1866: Meitzen, August (1868). *Der Boden und die landwirthschaftlichen Verhältnisse des Preussischen Staates*, Vol. 1-4. Berlin: Verlag von Paul Parey.

1871: Königliches Statistisches Bureau (1873-74). *Die Gemeinden und Gutsbezirke des Preussischen Staates und ihre Bevölkerung: Nach den Urmaterialien der allgemeinen Volkszählung vom 1. December 1871*, Vol. 1-11. Berlin: Verlag des Königlichen Statistischen Bureaus.

Königliches Statistisches Bureau in Berlin (1875). *Die Ergebnisse der Volkszählung und Volksbeschreibung im Preussischen Staate vom 1. December 1871*. Preussische Statistik Vol. 30. Berlin: Verlag des Königlichen Statistischen Bureaus.

1878: Herrfurth, Ludwig and Conrad Studt (1880). *Finanzstatistik der Kreise des preussischen Staates für das Jahr 1877/78*. Zeitschrift des Preussischen Statistischen Landesamtes, Ergänzungshefte, Vol. 7. Berlin: Verlag des Königlichen Statistischen Bureaus

1882: Königliches Statistisches Bureau in Berlin (1884/85). *Die Ergebnisse der Berufsstatistik vom 5. Juni 1882 im preussischen Staat*. Preussische Statistik Vol. 76 a-c. Berlin: Verlag des Königlichen Statistischen Bureaus.

1886 (Education): Königliches Statistisches Bureau in Berlin (1889). *Das gesammte Volksschulwesen im preußischen Staate im Jahre 1886*. Preussische Statistik Vol. 101. Berlin: Verlag des Königlichen Statistischen Bureaus.

1886 (Agriculture): Königliches Statistisches Bureau in Berlin (1887). *Die Ergebnisse der Ermittlung des Ernteertrags im preussischen Staate für das Jahr 1886*. Preussische Statistik Vol. 92. Berlin: Verlag des Königlichen Statistischen Bureaus.

1892: Neuhaus, Georg (1904). Die ortsüblichen Tagelöhne gewöhnlicher Tagearbeiter in Preußen 1892 und 1901, in *Zeitschrift des Königlich Preussischen Statistischen Bureaus*, Vol. 44, 310-346. Berlin: Verlag des Königlichen Statistischen Bureaus.

1896: Königliches Statistisches Bureau in Berlin (1897). *Die Ergebnisse der Ermittlung des Ernteertrags im preussischen Staate für das Jahr 1896*. Preussische Statistik Vol. 147. Berlin: Verlag des Königlichen Statistischen Bureaus.

1901: Neuhaus, Georg (1904). Die ortsüblichen Tagelöhne gewöhnlicher Tagearbeiter in Preußen 1892 und 1901, in *Zeitschrift des Königlich Preussischen Statistischen Bureaus*, Vol. 44, 310-346. Berlin: Verlag des Königlichen Statistischen Bureaus.

6. Existing Research using the iPEHD Data

A lot of research in economic history has used data from the iPEHD by now. This research is briefly described in this section. For those papers already published in academic journals, the iPEHD website provides ready-made datasets and codes in Stata to replicate the tables published in the papers. In addition, many more projects are currently under way and will be added to the website as publications become available. There is also a non-technical survey that summarizes some of the research conducted using the iPEHD data: Becker and Woessmann (2011a), “The Effects of the Protestant Reformation on Human Capital.”

6.1 Protestant Economic History and Education

Becker and Woessmann (2009), “Was Weber Wrong? A Human Capital Theory of Protestant Economic History” (started in 2006, first working-paper version released in 2007): This paper uses data from several censuses (Population 1871, Occupation 1882, Education 1886) and additional sources (including the Income Tax Statistics 1877) to show that the higher economic prosperity of Protestant relative to Catholic counties can be accounted for by

Protestants' higher literacy (presumably spurred by instruction in reading the Bible), suggesting that explanations based purely on differential work ethics may have limited explanatory power.

Becker and Woessmann (2008), "Luther and the Girls: Religious Denomination and the Female Education Gap in 19th Century Prussia": Using data from the first Prussian census in 1816, among others, this paper shows that a larger share of Protestants in a county's population decreased the gender gap in basic education.

Becker and Woessmann (2010), "The Effect of Protestantism on Education before the Industrialization: Evidence from 1816 Prussia" (first working-paper version released in 2009): This paper shows that Protestantism led to more schooling already in 1816, before the Industrial Revolution, ruling out that Protestant education just resulted from industrialization.

Becker and Woessmann (2011b), "Knocking on Heaven's Door? Protestantism and Suicide": Using data from 1816-21 and 1869-71, this paper finds a substantial positive effect of Protestantism on suicide.

6.2 Education and the Industrial Revolution

Becker, Hornung, and Woessmann (2011), "Education and Catch-up in the Industrial Revolution" (first working-paper version released in 2009): This paper combines school-enrollment and factory-employment data from 1816, 1849, and 1882 to show that – in contrast to the state-of-the-art view based on British evidence – basic education was significantly associated with non-textile industrialization in both phases of the Industrial Revolution.

Cinnirella and Hornung (2011), "Landownership Concentration and the Expansion of Education": Combining data from several censuses that effectively span the entire 19th century (1816, 1849, 1864, 1886, and 1896), as well as data from a 1866 classification of soil composition, this paper finds that landownership concentration, a proxy for the institution of serf labor, had a negative effect on school enrollment which diminished in the second half of the century.

6.3 Fertility and Education

Becker, Cinnirella, and Woessmann (2010), "The Trade-off between Fertility and Education: Evidence from before the Demographic Transition" (first working-paper version released in 2009): This paper uses data from the 1849 census and other sources to show that a

trade-off between child quantity and quality existed already in the 19th century and that causation between fertility and education runs both ways.

Becker, Cinnirella, and Woessmann (2012b), “The Effect of Investment in Children’s Education on Fertility in 1816 Prussia” (first working-paper version released in 2010): Using data from the 1816 census, this paper finds a significant negative causal effect of education on fertility – evidence for a child quantity-quality trade-off – already several decades before the demographic transition and shows that it is robust to accounting for spatial autocorrelation.

Becker, Cinnirella, and Woessmann (2012a), “Does Parental Education Affect Fertility? Evidence from Pre-Demographic Transition Prussia” (first working-paper version released in 2011): Combining data from three censuses – 1816, 1849, and 1867 – this paper finds a negative residual effect of women’s education on fertility, despite controlling for several demand and supply factors.

7. Additional Features of the iPEHD Website

The iPEHD website contains a number of additional features. For example, it provides a collection of thematic maps, produced using ArcGIS, that visualize the geographical distribution of several interesting variables across the Prussian territory. One such example is shown in Figure 1 above.

Furthermore, iPEHD is certainly not the only project dealing with historical Prussian data at the county level. Other projects provide such services as maps, information on territorial changes, additional data, and other material on Prussian counties. Links to websites of several of these projects, whose work is highly appreciated and can be viewed as complementary to iPEHD, are provided on the iPEHD website.

Finally, the iPEHD website contains a section on frequently asked questions, providing answers to standard problems encountered by iPEHD users.

8. Conclusions

The data contained in iPEHD, originally collected by the Royal Prussian Statistical Office, is an impressive collection of information whose quality, already in the 19th century, is generally regarded as excellent. Now digitized from censuses located in archives, these county-level data provide information on education, occupation, income and tax measures, industry, agriculture,

demographics, religion, and more. This database should facilitate future quantitative research on the economic history of 19th-century Prussia.

However, while iPEHD provides the service of supplying the historical data in a digitized and structured way and suggests ways on how to merge the data from different sources, researchers need to think carefully how to use the data in the context of their specific research projects. For instance, building panel datasets from the different census waves with varying administrative boundaries is a complex task that requires particular thought, meticulous care, and acquaintance with the structure of the original data. More generally, anybody planning to use the raw data contained in iPEHD should make sure to be well acquainted with the data structure and specifics as described in this documentation.

We hope that iPEHD provides a major service to researchers interested in Prussian economic history. Anybody who uses data from iPEHD is kindly requested to cite this paper as a source. Please also send one electronic copy of any work that uses data from iPEHD to iPEHD@ifo.de.

References

- Becker, Sascha O., Francesco Cinnirella, and Ludger Woessmann. 2010. "The trade-off between fertility and education: Evidence from before the demographic transition." *Journal of Economic Growth* 15, no. 3: 177-204.
- Becker, Sascha O., Francesco Cinnirella, and Ludger Woessmann. 2012a. "Does parental education affect fertility? Evidence from pre-demographic transition Prussia." *European Review of Economic History*: forthcoming.
- Becker, Sascha O., Francesco Cinnirella, and Ludger Woessmann. 2012b. "The effect of investment in children's education on fertility in 1816 Prussia." *Cliometrica* 6, no. 1: 29-44.
- Becker, Sascha O., Erik Hornung, and Ludger Woessmann. 2011. "Education and catch-up in the Industrial Revolution." *American Economic Journal: Macroeconomics* 3, no. 3: 92-126.
- Becker, Sascha O., and Ludger Woessmann. 2008. "Luther and the girls: Religious denomination and the female education gap in nineteenth-century Prussia." *Scandinavian Journal of Economics* 110, no. 4: 777-805.
- Becker, Sascha O., and Ludger Woessmann. 2009. "Was Weber wrong? A human capital theory of Protestant economic history." *Quarterly Journal of Economics* 124, no. 2: 531-596.
- Becker, Sascha O., and Ludger Woessmann. 2010. "The effect of Protestantism on education before the industrialization: Evidence from 1816 Prussia." *Economics Letters* 107, no. 2: 224-228.
- Becker, Sascha O., and Ludger Woessmann. 2011a. "The effects of the Protestant Reformation on human capital." In *The Oxford Handbook of the Economics of Religion*, edited by Rachel M. McCleary. Oxford: Oxford University Press: 93-110.
- Becker, Sascha O., and Ludger Woessmann. 2011b. "Knocking on heaven's door? Protestantism and suicide." CESifo Working Paper 3499. Munich: CESifo.
- Cinnirella, Francesco, and Erik Hornung. 2011. "Landownership concentration and the expansion of education." CESifo Working Paper 3603. Munich: CESifo.
- Galloway, Patrick R., Eugene A. Hammel, and Ronald D. Lee. 1994. "Fertility decline in Prussia, 1875-1910: A pooled cross-section time series analysis." *Population Studies* 48, no. 1: 135-158.